



**Access to State Government Information
Solutions Work Group Identification/Selection Committee**

Meeting: February 24, 2004

Present: Cheryl McLean
Swayzine McLean
Eileen Brown
Maury York
Kristin Martin

Jan Reagan
Lucy Reid
Tom Davis
Kelly Eubank

The Selection Committee discussed current practices of both the State Library and the State Archives and Records in terms of their selection process for collecting government information and concerns surrounding the identification and selection of digital information. The committee came up with definitions for record, publication, and webpage that include digital formats, and discussed selection criteria for selecting publications and other information to be part of a digital repository and a printed legacy collection. Finally, the meeting concluded with recommendations for next steps.

Current Practice: State Library

All agencies are required (executive, legislative, judicial, education institutions, boards and commissions) to send printed publications to the Clearinghouse. Each agency has one contact, usually the PIO. The Library keeps everything sent, although some items are excluded from the Clearinghouse (The *Handbook for State Agencies*, available at <http://statelibrary.dcr.state.nc.us/stdocs/handbook/aghndbk.htm> lists excluded items).

Monographs and serials are cataloged and added to the State Library collection; ephemera is distributed, but not cataloged or added to the permanent collection.

Current Practice: Records Center

Retention schedules for agency records identify which records should be saved permanently. A records analyst meets with agency staff to determine the schedule, but this does not happen as frequently as it should. Inactive records are transferred to the Records Center, but the agency still has intellectual control. Once records meet their retention, the 15% designated permanent are transferred to the archives, where they are further winnowed down and about 5% is preserved and made accessible to the public through the archives. Publications are listed on the schedules as going to the library.

Some Concerns and Issues Surrounding Identification and Selection

- Born digital information is not selected at all by the State Library and has no catalog record.
- The current practice for the Library is passive and many paper documents are not sent to the library. Staff turnover on both the library side and the agency side has been a problem. Additionally, changes in publishing practices have resulted in more decentralized publications, making it harder to track down publications.
- Current decisions to select and preserve digital objects and electronic records are being made by IT professionals, who think more about space considerations than the content of the information. Tom brought up the fact that Supreme Court briefs will not be kept as long as they should when considering research value, but instead will be deleted because of space issues.
- Current records policy requires agencies to print out electronic records, which they don't like to do, and sometimes doesn't make any sense.
- It's difficult to know what's out there.
- It's difficult to determine authenticity when digital information can be easily changed.
- You can't save everything – we need to bite off manageable chunks.

Definitions

Record: Everything's a record. Definition from G.S. 132: "Public record" or "public records" shall mean all documents, papers, letters, maps, books, photographs, films, sound recordings, magnetic or other tapes, electronic data- processing records, artifacts, or other documentary material, regardless of physical form or characteristics, made or received pursuant to law or ordinance in connection with the transaction of public business by any agency of North Carolina government or its subdivisions.

Publication: Information produced by state government designed for public dissemination. We may want to modify G.S. 125-11 definition to include digital publications. "State publication" means any document designed for public dissemination, including any report, directory, statistical compendium, bibliography, map, regulation, newsletter, pamphlet, brochure, periodical, bulletin, compilation, or register, regardless of physical format, prepared by a State agency or private organization, consultant, or research firm, under contract with or under the supervision of a State agency. Note: This keeps with the traditional publication definition from G.S. 125, but considers more formats than just print, and also dynamically generated items, like databases, that are the modern version of "statistical compendium."

Website: Both a publication and a record, a website consists of information made available on the web through hypertext transfer protocol or any other protocol. Websites are characterized by frequent changes to their content.

Collection Scope

The committee decided we should focus on information produced by state agencies in the 3 branches of government. From there we might move on to education institutions and local government information. We can develop guidelines to share with local government and encourage local libraries/institutions to collect that information. We felt we will need input from information facilitators, agency information creators, and the historical community when

selecting information. The State Library can lead the selection process or we could create an interagency taskforce to assist in selection.

Selection Criteria (criteria are based on Jimerson, Randall C. "Deciding What to Save." *OCLC Systems & Services* 19, no. 4 (2003): 135-40.):

1. **Functional Analysis:** the importance of the original purpose of the records, including the significance of the records creator, the creator's functions, and the records themselves.
Examples: information that documents the activities of state government, e.g. legal materials, statutes, court reports, ordinances, administrative code, governor's papers, etc.
2. **Content Analysis:** the informational significance of records, including importance of subjects represented in the records and the quality of documentation.
Examples: use of librarian judgment/experience to select, ways that information fits in with other information, political value, level of controversy, multiple language materials, use of the Records Center Website Content Assessment.
3. **Context Analysis:** the significance of the records in relation to other documentation, including consideration of duplication, uniqueness, format, scarcity, and record linkage.
Examples: publications within website environment, relationship between information in a records series.
4. **Use Analysis:** the value of records in meeting information needs and interests of the repository's clientele for past or projected uses, and the accessibility of the records and/or information.
Examples: publications most heavily used/mentioned in survey of depository librarians, hits of websites, other use statistics. Some concern that use now might not represent use in the future.
5. **Cost/Benefit Analysis:** weighing the value of information in the records against the cost of preservation, including staff time and facilities required for accessioning, processing, conservation, and long-term storage.
Examples: Spend more to preserve/provide access to most heavily requested items, consider the technical feasibility when selecting digital information.

Additional Discussion

- Not all information selected has to have the same level of preservation. Selection criteria can be used to determine which of the tiers of custodianship (levels of digital management) the materials should be preserved at.
- The website content assessment can be used to determine how frequently we should collect a snapshot of an agency's website.
- Committee members agreed with the broad selection criteria outlined in the proposed digital repository process. Digital versions of traditional publications would be selected for the highest level of preservation and access because they would receive library cataloging and be accessible through the catalog.

**North Carolina Department of Cultural Resources
Office of Archives and History**

WEBSITE CONTENT ASSESSMENT (WCA): Appraisal of the extent to which the agency is/is not fulfilling its legal/other responsibilities under records keeping statutes and other obligations, plus likelihood agency will be called to account. Outcome determines appropriate management and records keeping responses.

LOW LEVEL	MODERATE LEVEL	HIGH LEVEL
Characteristics: <ul style="list-style-type: none"> ▪ Single website ▪ Copies of official publications ▪ Controlled postings ▪ Publications not controversial ▪ Publications never litigated and little or no legal risk ▪ Publications do not generate bad press ▪ Constituents are satisfied w/ website ▪ Little legislative interest 	Characteristics: <ul style="list-style-type: none"> ▪ Number of complex websites ▪ Many offices posting to the websites ▪ Overlapping and conflicting information ▪ Both static and interactive websites ▪ Unique original materials on websites ▪ Legal counsel was not consulted ▪ Some adverse and controversial reaction ▪ Variety of management controls 	Characteristics: <ul style="list-style-type: none"> ▪ More numerous and complex websites w/ subsites ▪ Contain bulletin boards, publications, unique original materials, hearings, and real time business ▪ Different administrative arrangements ▪ Public scrutiny; publications are controversial ▪ Substantial liability; frequently litigated ▪ Frequent press coverage; public interest groups monitor sites ▪ Active and vigilant legislative interest ▪ Public is clamoring for more
Records Management Analysis: <ul style="list-style-type: none"> ▪ Satisfactory; no further measures necessary; periodic review of website for changes. 	Records Management Analysis: <ul style="list-style-type: none"> ▪ Additional measures should be taken to reconstruct prior versions 	Records Management Analysis: <ul style="list-style-type: none"> ▪ Requires precise reconstruction of exact copies of past contents
Records Keeping Response: <ul style="list-style-type: none"> ▪ Adequate documentation ▪ Records already in records keeping system (paper and electronic) ▪ Snapshot of entire website to Office of Archives and History at least every 2 years. 	Records Keeping Response: <ul style="list-style-type: none"> ▪ Maintain directory listings of websites ▪ Snapshot of entire website to Office of Archives and History at every significant version change or at least every 2 years, whichever occurs first ▪ Snapshots into the agency records keeping system and retention schedules applied 	Records Keeping Response: <ul style="list-style-type: none"> ▪ Maintain regular sequential directory listings of website changes ▪ Periodic snapshots of entire website (weekly, monthly) ▪ Snapshot of entire website to Office of Archives and History at every significant version change ▪ Snapshots into the agency records keeping system and retention schedules applied

Credit: McClure, Charles R. and J. Timothy Sprehe, Analysis and Development of Model Quality Guidelines for Electronic Records Management on State and Federal Websites, January 1998.

Legacy Collection

The committee agreed that having a core collection of documents preserved in paper is important. Determining what should be included in the legacy collection would be done using the same selection criteria described above. Publications that document the function of state government were immediately recognized as the most important items for the legacy collection. Some specific issues:

- When publications are no longer available in print the State Library will need to print out copies for its own collection. For other libraries that wanted a hard copy, the State Library could distribute microfiche.
- Publications that should be part of the legacy collection could be statutorily mandated to remain in print.
- Some “publications” might not be feasible to print out, particularly if they are databases or statistical data. The State Data Center could maintain the repository for digital statistical data, through LINC (Log Into North Carolina), which the SDC currently maintains.

Identification/Selection Committee Recommendations

1. Examine the current core collection to see what of these publications should form part of the legacy collection.
2. Research the statutes to find all publications defined by statute. Determine whether they should be part of the legacy collection.
3. Survey depository librarians regarding the most requested/used publications and their ideas for what should be part of a legacy collection.
4. Survey state agencies to identify their most important/used publications and what they think should be part of the legacy collection. Also work with state agency webmasters to find out which websites are the most popular, such as an analysis of website hits.
5. Talk to the State Data Center about what they see as the most important statistical data. and whether they could serve as the repository for the legacy collection of digital data.
6. Create a list of all of the agency domain websites.
7. While not definite, the State Library may be part of a grant from the Library of Congress National Digital Information Infrastructure and Preservation Program to develop a semi-automated selection tool for digital information. This could fit in with the criteria developed through this committee.